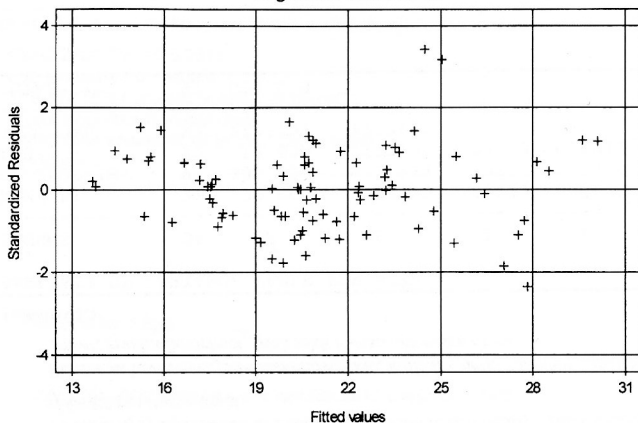


Kuva alla esittää estimoidun mallin (4.1) standardoituja residuaaleja:

Regression Residual Plot



Residuaaleihin on sovitettu *apuregressio*

$$(4.2) \quad e_j^2 = \alpha_0 + \alpha_1 \hat{y}_j + \delta_j$$

jossa

e_j = estimoidun mallin residuaali

\hat{y}_j = estimoidun mallin sovite

Estimointitulokset *apuregressiosta* (4.2) on annettu seuraavalla sivulla.

STATISTIX FOR WINDOWS					CITYDAT
UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF RESSQR					
PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P	
-----	-----	-----	-----	-----	
CONSTANT	-15.0943	11.9630	-1.26	0.2104	
FIT	1.23697	0.56043	2.21	0.0299	
R-SQUARED	0.0525	RESID. MEAN SQUARE (MSE)	378.099		
ADJUSTED R-SQUARED	0.0417	STANDARD DEVIATION	19.4448		
SOURCE	DF	SS	MS	F	P
-----	-----	-----	-----	-----	-----
REGRESSION	1	1841.99	1841.99	4.87	0.0299
RESIDUAL	88	33272.7	378.099		
TOTAL	89	35114.7			
CASES INCLUDED 90					MISSING CASES 0

Apuregression (4.2) selitysteesta R^2 laskettiin testisuure

$$(4.3) \quad nR^2 = 90 \times 0.0525 = 4.725$$

jossa on n on havaintojen lukumäärä.

Erään nollahypoteesin pätiessä

$$nR^2 \sim \chi^2(1).$$

Testisuureen (4.3) arvoa 4.725 vastaava p-arvo on 0.02973.

Tehtävät:

- Ovatko kaikki mallin (4.1) regressiokertoimet merkitseviä 1 %:n merkitsevyystasolla?
- Mikä on estimoidun mallin (4.1) selityste? Mitä johtopäätöksiä voit tehdä tulostuksen F -testistä?
- Mikä on suureiden R-SQUARED ja ADJUSTED R-SQUARED ero?
- Onko multikollineaarisuus ollut estimoinnissa ongelma?
- Miksi alkuperäisen regressiomallin (4.1) residuaaleihin on sovitettu *apuregressio* (4.2)?
- Mitä nollahypoteesia testisuurella (4.3) on testattu? Mikä on testin tulos?